LATENT TOPIC ANALYSIS METHODS TO MONITOR HEALTH ON SOCIAL MEDIA

B Neelima #1, N Padmavathi #2, V V Naga Pranrtha #3, S Sreelatha #4, M Bala Krishna Reddy #5, P Anusha #6
#1Asst. Professor, #2,3,4,5,6 B.Tech.., Scholars
Department of Computer Science and Engineering,
QIS College of Engineering and Technology

Abstract:

The use of social media as a primary resource for research on any and all elements of everyday life is becoming more common. Twitter now allows for specialised latent topic analysis approaches, such as the Ailment Topic Aspect Model (ATAM), which makes it possible to monitor public health conversations. The use of social media to track changes in people's health over time is of particular interest to us in this line of study. The usage of tweets comes with a number of advantages, the most notable of which is the availability of data almost instantly at no cost. Early monitoring of health data is a useful adjunct to post-mortem examinations and paves the way for a variety of applications, including the assessment of behavioural risk factors and the initiation of health awareness programmes. We pose two problems: one is the detection of a health transition, and the other is the prediction of a health transition. First, we present the Temporal Ailment Topic Aspect Model, often known as TM–ATAM. This is a brand new latent model that has been designed specifically to solve the first challenge by capturing transitions that include topics that are connected to health. ATAM has been extended in a way that isn't immediately clear in order to create TM–ATAM, which was made to extract information on medical-related issues. It does this by reducing the prediction error on topic distributions across subsequent postings at various temporal and geographic granularities, and then using this information to learn health-related topic transitions. In order to address the second issue, we devised T–ATAM, which stands for Temporal Ailment Topic Aspect Model. Within this model, time is seen as a random variable on a fundamental level.

1. **Introduction**:

The life that is based on the internet has become an important source of information that can be used to dissect all aspects of regular living.

Twitter is specifically used for general health observation in order to decipher early signs of the prosperity of populations in a variety of geographic locations. The ability to model transitions for diseases and observe statements such as "people concerning smoking and cigarettes before talking about metastasis problems" or "people cite headaches and abdomen pain in any order" is beneficial for syndromic surveillance in the health domain. It also helps live activity risk factors and triggers public health campaigns. In this article, we will be calculating two different concerns. The identification of health transitions and the prediction of health transitions are both important. We have a tendency to construct TM–ATAM that predicts the temporal transitions of health-related subjects in order to mitigate the negative effects of discovery. T–ATAM is a novel method that treats time as a variable natively inside ATAM [4, which enables it to identify latent disorder within tweets]. This is done so as a means of overcoming the challenge posed by the problem of prediction.

When trying to forecast the subtle shifts that occur in the health-related debate on Twitter, using time as a variable is absolutely necessary.

Sentinel observation is the method that is used to collect data from hospitals and other medical institutions in order to monitor common diseases. The task of law enforcement is hampered by such assets, particularly with regard to ongoing criticism. The internet has evolved into a source of syndromic surveillance that can be carried out on a larger scale, in close proximity to real time, and for very little cost. Our problems are as follows: I identifying tweets that are connected to health; (ii) determining when health-related conversations on Twitter shift from one issue to another; and (iii) capturing distinct transitions of this kind in different geographical areas. In point of fact, not only do disorder distributions change throughout the course of time, but they also change geographically. Because of this, in order for us to be successful, we need to carefully represent the two most important granularities, which are temporal and spatial.

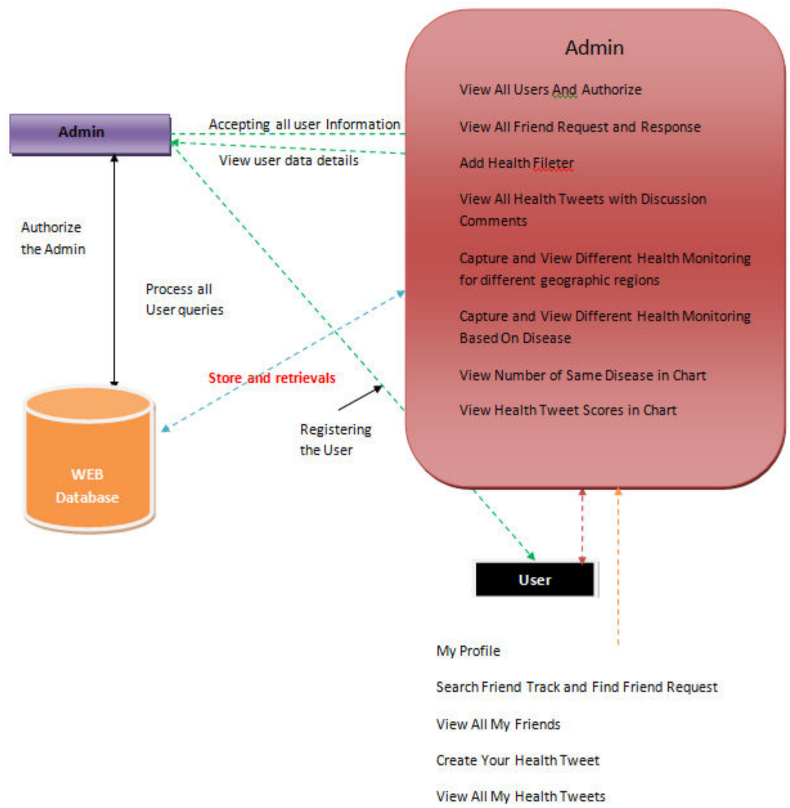
On a larger scale, in close to real time, and at almost no additional expense. Our problems are as follows: I identifying tweets that are relevant to health; (ii) determining when health-related conversations on Twitter shift from one subject to another; and (iii) capturing distinct transitions of this kind for various geographical locations. In point of fact, not only do disease prevalence rates change over the course of time, but they also change geographically.

Because of this, in order to achieve efficiency, we need to precisely represent not one but two fundamental granularities: temporal and spatial. A temporal granularity that is too small may lead to sparse and false transitions, whereas a temporal granularity that is too coarse may cause the missing of important illness changes.

In a similar vein, a geographic granularity that is too fine may yield false positives, while a geographic granularity that is too coarse may miss crucial transitions, such as when it comes to people living in varying climates. For instance, talks on allergies take a pause during various times of the year and in various locations throughout the United States [4]. Because of this, if you analyse all tweets coming from the United States collectively, you will overlook climatic fluctuations that effect people's health. In this paper, we propose that it is necessary to take into account distinct timegranularities for the various areas, and we want to discover and predict the development of disease distributions across different temporal granularities.

It has been shown that dedicated methods such as the AilmentTopic Aspect Model (ATAM) are better suited for capturing ailments in Twitter [4]. Although it has been proposed that several latent topic modelling methods, such as Probabilistic Latent Semantic Indexing (pLSI) [5] and LatentDirichlet Allocation (LDA) [6], can effectively cluster and classify general-purpose text, it has also been shown that these methods are not the only ones that can do so.

ATAM is an extension of LDA that models the manner in which people express symptoms in tweets. It assumes that every health-related tweet indicates an underlying condition such as influenza or allergies. An illness acts as an indicator of word distribution in the same way that a subject does. In addition to this, ATAM keeps a distribution across treatments and symptoms. This degree of specificity results in a model that is more accurate for underlying conditions.



2. System Analysis

The authors of the existing system present a technique that learns the shifting word distributions of subjects over time. Additionally, the writers of the existing system utilise the structure of a social network to learn how topics temporally change in a community. However, TM–ATAM and T–ATAM are not the same as dynamic topic models like [9] and [10], as well as the work of

Wang et al. [11], because they are intended to learn topic transition patterns from temporally-ordered posts, whereas dynamic topic models concentrate on the shifting word distributions of topics over the course of time.

TM–ATAM is able to learn transition parameters that determine the progression of health-related subjects by reducing the prediction error on illness distributions throughout successive periods at varying levels of temporal and geographic granularity. This process takes place during the learning phase. T–ATAM, on the other hand, analyses health tweets in such a way that it treats time as a corpus-specific multinomial distribution. This allows it to uncover previously unknown illnesses.

The mining of themes for the purpose of deducing citations has been done using more traditional methods. In order to conduct an empirical investigation on topic modelling and time-based topic modelling, respectively, many other discriminative methodologies have been used. None of these may be immediately applied to the facts about health.

Disadvantages

- There is currently no way to map documents to tweets.
- There is a feature on ATAM called "Uncovering Health Topics."

3. Proposed System

The health transition detection issue and the health transition prediction problem are formulated by the system in the proposed system, and the system then solves both of these problems. In order to solve the detection issue, the system creates TM–ATAM, which mimics the temporal transitions of several health-related subjects. T–ATAM is a revolutionary approach that reveals latent disease inside tweets. This is accomplished by considering time as a random variable natively within ATAM. In order to solve the challenge of prediction, we have proposed T–ATAM as a solution.

The key to accurately anticipating the gradual shift in the tone of health-related conversation on Twitter is to treat time as a random variable.

Advantages

- TM–ATAM is an algorithm that monitors tweets linked to health and tracks how those tweets change over time and location. TM–ATAM will learn the transition parameters for a specific area by reducing the prediction error for illness distributions across predetermined time periods.

- T–ATAM is an innovative new model that is able to predict tweets that are connected to health by using time as a variable whose values are derived from a multinomial distribution that is peculiar to the corpus.
- Extensive trials that indicate that T–ATAM is better than TM–LDA and TM–ATAM for forecasting health transitions, when contrasted against its efficacy against a ground truth.

4. Implementation

Admin

Within this section, the Admin is required to log in by providing a user name and password that are both legitimate. After a successful login, he will be able to carry out a number of tasks, including View All Users And Authorize. Explore Every Friend Request and Its Response, Add Health Filter, View All Health Tweets with Discussion Comments, Capture and View Different Health Monitoring for Different Geographic Regions, Capture and View Different Health Monitoring Based On Disease, Add Health Filter, View the Number of Individual Diseases in the Chart and View the Scores of Health Tweets in the Chart

Friend Request & Response

Within this section, the admin will be able to examine all of the friend requests as well as the answers. In this section, all of the requests and answers will be presented along with their respective tags, which may include ID, requested user picture, requested user name, user name request to, status, time and date. If the user does not accept the request, the status will stay as pending; however, if the user does accept the request, the status will be changed to accepted.

User In this particular module, there are a total of n users currently logged in. Before carrying out any actions, the user is required to register. The information of the user will be saved to the database as soon as they have registered. After his successful registration, he will be required to log in using the user name and password that he was assigned. Verify finger print and Login Once the user has successfully logged in, they will have access to a number of features and functions, including My Profile, Search Friend Track and Find Friend Request, View All My Friends, Create Your Health Tweet, View All My Health Tweets, View and Monitor All My Friends Health Tweets, and View All My Health Tweets.

Looking for other users to make friends with

The user will search for other users in the Networks as well as in the Same Network in this module, and then they will send friend invitations to those persons. If the user has authorization, they are able to seek for other users on different networks in an effort to establish new friends.

5. Conclusion

We design strategies to extract information about illnesses from social media over time. We outlined challenges involving the detection and prediction of health transitions, and then we offered two models to handle those difficulties. The TM–ATAM model, which is based on granularity and is used to perform area-specific analysis and ultimately leads to the identification of temporal periods and the characterization of homogenous disease discourse per region, is what is used to address detection. The T–ATAM model, which interprets time in its original form as a random variable whose values are chosen from a multinomial distribution, is used to solve the problem of prediction. Due to the fine-grained nature of T–ATAM, the modelling and prediction of transitions in tweets linked to health do not significantly improve. Our methodology, as far as we can tell, is relevant to a variety of different fields, including those dealing with time-sensitive issues like emergency management and national security concerns.

References

[1] L. Manikonda and M. D. Choudhury, “Modeling and understanding visual attributes of mental health disclosures in social media,”in Proceedings of the 2017 CHI Conference on Human Factors inComputing Systems, Denver, CO, USA, May 06-11, 2017., 2017, pp.170–181.

[2] S. R. Chowdhury, M. Imran, M. R. Asghar, S. Amer-Yahia, andC. Castillo, “Tweet4act: Using incident-specific profiles for classifyingcrisis-related messages,” in 10th Proceedings of the InternationalConference on Information Systems for Crisis Response andManagement, Baden-Baden, Germany, May 12-15, 2013., 2013.

[3] T. Davidson, D. Warmsley, M. W. Macy, and I. Weber, “Automatedhate speech detection and the problem of offensive language,” inProceedings of the Eleventh International Conference on Web and SocialMedia, ICWSM 2017, Montréal, Québec, Canada, May 15-18, 2017.,2017, pp. 512–515.

[4] M. J. Paul and M. Dredze, “You Are What You Tweet: AnalyzingTwitter for Public Health,” in ICWSM’11, 2011.

[5] T. Hofmann, “Probabilistic Latent Semantic Indexing,” in SIGIR’99, 1999, pp. 50–57.

[6] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent DirichletAllocation,”Journal of Machine Learning, vol. 3, pp. 993–1022, 2003.

[7] Y. Wang, E. Agichtein, and M. Benzi, “TM-LDA: Efficient OnlineModeling of Latent Topic Transitions in Social Media,” in KDD’12,2012, pp. 123–131.

[8] S. Sidana, S. Mishra, S. Amer-Yahia, M. Clausel, and M. Amini,“Health monitoring on social media over time,” in Proceedingsof the 39th International ACM SIGIR conference on Research andDevelopment in Information Retrieval, SIGIR 2016, Pisa, Italy, July17-21, 2016, 2016, pp. 849–852.

[9] D. M. Blei and J. D. Lafferty, “Dynamic Topic Models,” in ICML’06,2006, pp. 113–120.

[10] C. X. Lin, Q. Mei, J. Han, Y. Jiang, and M. Danilevsky, “The JointInference of Topic Diffusion and Evolution in Social Communities,”in ICDM’11, 2011, pp. 378–387.

[11] X. Wang and A. McCallum, “Topics Over Time: A Non-MarkovContinuous-time Model of Topical Trends,” in KDD’06, 2006, pp.424–433.

[12] K. W. Prier, M. S. Smith, C. Giraud-Carrier, and C. L. Hanson,“Identifying Health-related Topics On Twitter,” in Social computing,behavioral-cultural modeling and prediction. Springer, 2011, pp.18–25.

[13] C. Cortes and V. Vapnik, “Support-vector networks,” MachineLearning, vol. 20, no. 3, pp. 273–297, 1995. [Online]. Available: http://dx.doi.org/10.1007/BF00994018

[14] M. De Choudhury, “Anorexia on Tumblr: A CharacterizationStudy,” in DH’15, 2015, pp. 43–50.

[15] M. De Choudhury, A. Monroy-Hernández, and G. Mark, “"narco"Emotions: Affect and Desensitization in Social Media During theMexican Drug War,” in CHI’14, 2014, pp. 3563–3572.

[16] U. Pavalanathan and M. De Choudhury, “Identity Managementand Mental Health Discourse in Social Media,” in WWW’15, 2015,pp. 315–321.

[17] F. Bouillot, P. Poncelet, M. Roche, D. Ienco, E. Bigdeli, andS. Matwin, “French Presidential Elections: What are the MostEfficient Measures for Tweets?” in PLEAD’12. ACM, 2012, pp.23–30.

[18] L. Hemphill and A. J. Roback, “Tweet Acts: How ConstituentsLobby Congress via Twitter,” in CSCW’14, 2014, pp. 1200–1210.

[19] A. Ceron, L. Curini, and S. M. Iacus, “Using Sentiment Analysis toMonitor Electoral Campaigns: Method Matters-Evidence from theUnited States and Italy,” Soc. Sci. Comput. Rev., vol. 33, no. 1, pp.3–20, 2015.

[20] P. Barberá, "Birds of The Same Feather Tweet Together: BayesianIdeal Point Estimation using Twitter Data," Political Analysis,vol. 23, no. 1, pp. 76–91, 2015.

[21] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependentTwitter Sentiment Classification," in HLT'11, 2011, pp. 151–160.